Supek & Lehner (2019) *DNA Repair*, 81:102647. https://doi.org/10.1016/j.dnarep.2019.102647

Scales and mechanisms of somatic mutation rate variation across the human genome

Fran Supek [1,2] *, Ben Lehner [2,3,4]

[1] Institut de Recerca Biomedica (IRB Barcelona), The Barcelona Institute of Science and Technology, 08028, Barcelona, Spain.
[2] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain.
[3] Systems Biology Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Doctor Aiguader 88, 08003 Barcelona, Spain.
[4] Universitat Pompeu Fabra (UPF), Barcelona, Spain.
* e-mail: fran.supek@irbbarcelona.org

Abstract

Cancer genome sequencing has revealed that somatic mutation rates vary substantially across the human genome and at scales from megabase-sized domains to individual nucleotides. Here we review recent work that has both revealed the major mutation biases that operate across the genome and the molecular mechanisms that cause them. The default mutation rate landscape in mammalian genomes results in active genes having low mutation rates because of a combination of factors that increase DNA repair: early DNA replication, transcription, active chromatin modifications and accessible chromatin. Therefore, either an increase in the global mutation rate or a redistribution of mutations from inactive to active DNA can increase the rate at which consequential mutations are acquired in active genes. Several environmental carcinogens and intrinsic mechanisms operating in tumor cells likely cause cancer by this second mechanism: by specifically increasing the mutation rate in active regions of the genome.

Abbreviations

CTCF, CCCTC-binding factor; H3K36me3, histone H3 trimethylated at lysine 36; DHS, DNase I hypersensitive sites; RT, replication timing; MMR, mismatch repair; UV, ultraviolet light; UTR, untranslated region; NER, nucleotide excision repair; ETS, E-twenty-six protein family; TF, transcription factor; BER, base excision repair; AID, activation-induced cytidine deaminase

Keywords

mutation rates; somatic cells; tumors; genomic instability; DNA mismatch repair; chromatin

Introduction

The large-scale sequencing of tumors and healthy somatic cells presents a unique opportunity to learn about somatic mutation processes and how mutation rates vary across the human genome. The primary motivation for tumor genome sequencing was to identify the 'driver' mutations that cause cancer. Driver mutations are selected for because they promote the expansion or survival of tumor clones. However, most somatic mutations in cancer genomes are inconsequential 'passenger' mutations that are under very weak or no selection and statistical analyses of these passenger mutations have provided many fundamental insights into the mutation processes that operate in human cells and how these processes vary across the genome, cell types and individuals.

Absolute mutation rates are difficult to determine for tumor cells, primarily because the number of cell divisions that a tumor cell has undergone is hard to establish. However, it has long been appreciated that many tumors have have an elevated mutation rate, for example because of inactivated DNA repair pathways [1–3]. In this review, we will not focus on the general acceleration in mutation rates in a cancer cell. Instead, this text focuses on relative mutation rates, which are more straightforward to quantify from regional densities of mutations in the genome. We provide an overview of the patterns of mutations that are observed across regions of the human genome and our current understanding about their mechanistic underpinnings when this is known (although often a detailed mechanistic understanding is still lacking). We place an emphasis on the insight and the novel hypotheses that cancer genomes have yielded about the organization of mutation processes in human cells. Our primary focus is on single nucleotide substitutions and short insertions and deletions. The reasons for this are pragmatic: structural variation is much more challenging to precisely infer using short-read sequencing and although progress is being made in both identifying and understanding structural variation, the influences on its regional rates in the soma are far less well understood [4].

Variability in mutation rates across the genome may result from two broad causes: differential accrual of DNA damage and also base mispairing during DNA replication (variation in mutation supply) or differential repair of damage and mispairs (variation in DNA repair). These influences are, of course, not mutually exclusive. Recent work has suggested, however, that the latter – differential DNA repair – appears to play a quantitatively more important role in shaping the mutation landscape in the human soma. This is consistent with the expectation that mutation

rates are more sensitive to changes in repair rates than to changes in damage rates because the vast majority of instances of damage are repaired [5].

Somatic mutation rates vary at multiple resolutions

As we discuss below, mutation rates in the human genome vary at multiple different scales from single nucleotides to megabase-sized domains. Importantly, the mechanisms underlying variation at these different genomic resolutions may be quite different and this often confounds statistical analyses performed at a certain resolution. For instance, at the resolution of a single-nucleotide, mutation rates are highly dependant on the 5' and 3' neighboring nucleotides. For example, in the human genome, the spontaneous deamination of methylated cytosine to thymine results in a substantially increased mutation rate at CG dinucleotides, the majority of which are methylated in the genome [6,7]. At the other extreme, when examining roughly megabase-sized chromosomal domains, a major determinant of mutation rates is DNA replication timing. In this specific case, this is not due to differential damage accumulation, but due to differential activity of DNA repair, which is preferentially active in early-replicating, gene-rich domains [8]. We will discuss these and other known determinants of regional mutation rates at length below.

Genome-wide statistical analyses of the patterns in mutation rate heterogeneity have revealed mechanisms that marshall DNA repair preferentially towards particular regions of the genome. Given that the cell's ability to repair DNA is a limited resource, it is not unexpected that different repair capacities would be prioritized to different regions, for example those containing genes essential for cell viability. This differential DNA repair controls the mutation supply to oncogenes and tumor suppressor genes and may therefore modulate the risk of cancer. Potentially, the mutation supply to important genes in somatic cells could more broadly affect aging-related diseases, such as neurodegeneration [9]. With respect to germline mutations, variation in mutation rates along the chromosomes determines the propensity for obtaining particular deleterious mutations, and also shapes the genetic diversity of populations. Learning about patterns of regional mutation rates also has important practical applications. One common use concerns statistical tests for selection in genomic sequences; such tests are highly dependant on establishing an accurate baseline for local mutational rates, and any deviations therefrom are then considered as evidence for either positive selection (in cancer, suggesting oncogenic mutations) or negative selection (suggesting genes essential to somatic cells).

In between the two extremes mentioned above – the trinucleotide scale and the megabase domain scale – there is a continuum of resolutions of mutation rate variability which have been less systematically explored. However there is knowledge that rates of certain mutational processes vary depending, for instance, on the binding of certain proteins (most prominently, the sites where the CTCF protein is co-bound with cohesin [10–12]) or on the presence of certain post-translational histone modifications on nucleosomes – a salient example is the H3K36me3 mark associated with transcription elongation that can recruit DNA repair proteins [13–15]. Interactions exist between the mutation patterns at different resolutions, which helps link mutation patterns to mechanisms. For example, the CTCF-bound motifs were found to be more mutable, in relative terms, in the same tumors which have a higher proportion of A>C changes, which allows these two processes to be tentatively linked to the same underlying mechanism, suggested to be oxidation of the free nucleotide pool [11,16]. Such associations may help narrow down a list of putative mechanisms and thus to prioritize experimental work.

In summary, what is known is that regional mutation rates in somatic cells appear highly variable at different, overlapping scales. What is currently less well explored is how much systematic variability there exists at each resolution (while rigorously deconvoluting the variability that overlaps and being able to distinguish it from random noise), how such variability at every resolution changes between cell types and between individuals, and how much of the variability is due to differential DNA repair and how much is due to differential damage. More fundamentally, the identity of the repair and/or damage mechanisms that cause the observed patterns are still often unknown.

Features and mechanisms associated with mutation rate variation

In the following sections we provide an overview of genomic and epigenomic variables known to be statistically associated with mutation rates at various resolutions. We also highlight examples where the underlying molecular mechanisms are known or suspected.

Variability at the domain scale (10^5 bp - 10^6 bp). The extent of variability in somatic mutation rates at the scale of roughly megabase-sized domains was appreciated early on, after the sequencing of the first cancer genomes [17–19]. High mutation rates are strongly correlated with a high density of repressive chromatin marks (such as the heterochromatin mark

H3K9me3), later replication timing, lower accessibility of DNA as estimated via the density of DNAse hypersensitive sites (DHS) in the domain, and a lower density of active chromatin marks (e.g. H3K4me1/2/3). However, correlation does not imply causation and indeed for several of these variables there is mounting evidence they are likely not causal to mutation rates, or at least not to a major degree. Based on statistical analyses that rigorously control for confounding variables, replication timing (RT) is a feature that is very robustly related with mutation rates [15,19–21]. Consistently, changes in RT across cell types correlate well with changes in mutation rates [8]. The key test of the hypothesis would, however, be an experiment that changes RT in a controlled manner whilst measuring the impact on additional features such as chromatin modifications, and this has not yet been performed.

The converse example of a mutation rate determinant, often cited as associated in the literature, is chromatin accessibility (usually measured by the density of the DHS), which is associated with locally lower mutation rates [22]. However, meticulous statistical analyses suggest that DNA accessibility is less likely to be causal for many types of mutational processes, or at least the processes commonly generating a high number of mutations [15]. An informal interpretation of these analyses is that the large, usually several-fold difference in mutation rates between chromosomal domains cannot be explained by the cumulative effect of smaller differences in mutation rates between accessible sites (DHS) and their local neighborhoods in DNA. In other words, the density of DHS sites at the megabase scale, strongly correlated with mutation rates, likely reflects a different mechanistic cause because at the local scale DHS have a more subtle effect on mutability. (In fact, the density of DHS site was found to be an accurate predictor of RT [23], thus explaining the confounding.) Similar reasoning would hold for the strong correlation of the domain-scale density of the enhancer/promoter chromatin marks (H3K4 methylation) [24], which however at the local scale show little association with mutability. They are therefore unlikely to be causal of the domain-scale variation in mutation rates.

There is mounting evidence for a mechanism underlying the striking domain-scale variability in mutation rates: differential activity of DNA mismatch repair (MMR), which preferentially prevents mutations in early-replicating, euchromatic regions. This was first discovered because MMR-deficient tumors of various organs display a severe loss of domain-scale variability, exhibiting a ‘flat’ landscape [8], a finding later confirmed by direct experiments [25,26]. Further evidence came from analysis of trinucleotide mutation spectra: those 5’ and 3’ contexts which become more mutable upon MMR failure are the same contexts that become less depleted in early-

replicating DNA [8,27,28]. The importance of MMR for domain-scale mutation rate variability has also been confirmed by direct experiments where MMR activity was abolished in human cell lines [25,26] resulting in a 'flat' distribution of mutations across the genome. Similar results have also been reported for MMR in yeast [5,29] and in *Arabidopsis* [30]. Additionally, nucleotide excision repair (NER) appears to be less active in late-replicating heterochromatin, because the mutation rate differential is diminished in skin cancers of germline NER-deficient patients [31]. Differential DNA repair therefore seems to be a major reason that mutability landscapes exist, both in the human genome and in the other eukaryotes.

There are several open questions related to this domain-scale mutation rate variation. First, does MMR decrease in accuracy, or in efficiency, or both in late-replicating DNA and what is the mechanism causing this? Plausible mechanisms could be the preferential loading of MMR complexes onto early replicating euchromatin [13], depletion or modification of a repair factor late in S-phase, the reduced accessibility of heterochromatin to repair factors, or simply the reduced time available for the repair of late replicating DNA [32,33]. The second open question concerns the mechanism by which MMR reduces various mutation types. Mismatch errors in DNA replication are a natural substrate for MMR and likely to be one of the most abundant mutation types. However certain types of mutations that are associated, for examples, with bulky nucleotide adducts or UV-related DNA damage, also appear to have similar domain-scale distribution in mutation rates in tumor genomes and it is not clear how MMR contributes to their repair. One possible explanation is that MMR proteins can serve as sensors for damage, binding to a lesion to promote recruitment of other repair pathways [34,35]. Further analyses and experimental work are needed to address these questions.

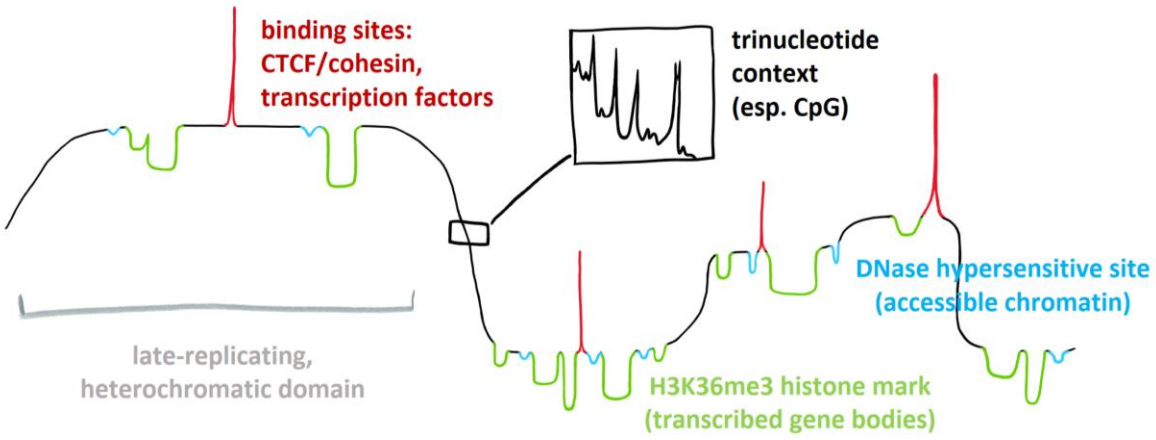

Figure 1. Variation in regional mutational rates in human somatic cells at multiple overlapping scales.

Variability at the gene scale (10^3 bp - 10^5 bp). Cancer genomics has provided ample evidence that mutation rates differ between different genes in a manner independent of differences between chromosomal domains. Several types of statistical associations were reported, sometimes with suggestive evidence for mechanisms. The effect size of such between-gene differences in mutation rates is usually more modest than the striking, several-fold differences between chromosomal domains. Nonetheless such mutation rate gradients at the genes (meaning: between genes, and, in some instances, across the parts within the gene body) are of interest because they are informative about mechanisms of mutagenesis and DNA repair. An additional interest lies in the fact that mutagenesis within genes -- particularly coding regions, UTRs, and promoters -- is more likely to have downstream functional consequences for cancer, aging and other diseases.

The most striking trend in mutation patterns observed at the gene-scale is transcription-related and affects the entire gene body: the asymmetry in mutation rates between the transcribed and the non-transcribed DNA strands. This is evident only for some mutational processes, falling into two groups: a strand bias resulting from DNA damage that can be seen by the transcription-coupled nucleotide excision repair (NER) pathway where the strand bias results from differential NER [36–38] and a strand bias resulting from transcription-coupled damage, which is mechanistically currently mysterious and was reported in liver tumors [39] and in healthy neurons [40]. Most mutational processes (as defined by the trinucleotide mutation signatures) do not exhibit a transcriptional strand bias [36,39].

In addition, higher transcription levels are generally associated with reduced mutation rates in both strands of genes [37,38], although it is not clear to which extent this is due to the correlation of expression levels with other genomic features, such as the abundance of higher-expressed genes in early replication time regions. Some very specific cases of mutational processes may also have elevated rates in highly transcribed regions but without clear evidence of a mutational strand bias, for instance transcription-associated mutagenesis due to oxidative damage was reported in a human cell line [34] and highly expressed lineage-specific genes in several tumor types harbored many indel mutations [41].

One widespread mechanism by which highly expressed genes have a reduced mutation rate is revealed by the association between a particular histone modification – H3K36me3 – and increased DNA repair efficiency. H3K36me3 is a mark associated with transcription elongation, deposited by a histone methyltransferase recruited by the elongating form of the RNA polymerase II. H3K36me3 can recruit the MSH6 protein of the MMR pathway [13] and in a large-scale analysis of cancer genomes the high levels of the H3K36me3 mark were found to have a strong association with up to two-fold lower mutation rates, after rigorously controlling for other confounding variables such as DNA accessibility and replication time, thus suggesting a causal role of H3K36me3 [15]. This association disappears in MMR-deficient cancers, providing evidence that MMR recruitment causes lower mutation rates at H3K36me3-marked regions [15], which predominantly occur in bodies of highly transcribed genes and increase from the 5' towards the 3' gene end [42,43]. Experimental work in a cell line model provides strong support for the causal role of H3K36me3 in protection from mutations via marshalling MMR activity towards expressed genes [34]. Of note, the H3K36me3 mark has also been proposed to recruit another type of DNA repair – homologous recombination [14].

Within some genes, exons may accumulate slightly higher levels of H3K36me3 than introns, which is consistent with modestly reduced mutation rates in exons observed in cancer [22,34,44], which might thus result from differential repair. However this observation is confounded by technical issues: the repeat content and G+C content differences between exons and introns may affect DNA sequencing, short read alignment and mutation calling; upon stringent filtering of mutations and control for confounding mutational processes, there were no observable differences in mutability between exons and introns [45]. Additionally, negative selection against exonic splicing elements has been suggested to reduce exonic mutation density [46]. A further observation involves an apparent depletion of oxidative DNA damage in exons compared to introns [47]. These associations and proposed mechanisms are not necessarily mutually exclusive and need to be resolved. The modest effect size complicates statistical analyses to deconvolute these factors.

In addition to the mechanisms above, which for the most part invoke differential DNA repair to explain lower mutation rates in active chromatin, there is experimental evidence that differential DNA damage might, at least in part, underlie lower mutation rates in active regions. For instance, genome-wide maps of UV damage suggest that this type of insult affects inactive,

heterochromatic regions more strongly [48], potentially because of the peripheral placement of heterochromatin in the nucleus. In a similar vein, genic regions were reported to have a lower rate of oxidative damage compared to intergenic regions [47]. Broadly, this distribution of damage does indeed mirror the distribution of mutations observed in somatic cells undergoing UV or oxidative-stress related mutagenesis [47,48]. Further experimental work that combines both differential damage mapping and mutation rate heterogeneity data in a rigorous joint analysis is warranted to quantitate the contributions of differential damage versus differential repair to individual mutational processes.

Variability at the sub-gene scale (10^1 bp-10^3 bp). Various genomic regions considerably smaller than the size of genes have also been associated with certain types of mutational processes in certain cell types: mutational processes operating at this scale tend to have higher tissue- and mutagen-specificity than processes at larger scales. This is perhaps unsurprising, given that such processes (at least the ones described thus far) often result from binding of a particular protein to a certain site in DNA, rather than a global property of the genome maintenance machinery.

A salient example of such phenomena is the high mutation rate at the binding sites of the CTCF protein, a regulator of transcription and of chromatin architecture. CTCF sites were found to be highly mutable in colon cancer, stomach cancer, liver cancer and melanoma [10–12,49], where hypermutation is conditional on the concomitant binding of cohesin, a partner protein, to the site. There is suggestive data that this may be due to exclusion of DNA repair processes from the CTCF sites, including both MMR and NER [11,12]. The exact sites in (or adjacent to) the CTCF motif that are hypermutable, intriguingly, differ between the mutagenic exposures: UV mutagenesis is associated with G>A/C>T changes at one set of sites in the motif [12], while mutagenesis due to ‘Signature 17’ (putatively, oxidative damage to the guanine in the free nucleotide pool [16]) is associated with A>C/T>G mutations at a different set of sites [11]. Overall, therefore, both differential DNA damage and differential repair are likely to play a role in CTCF site hypermutability. An open question is certainly whether such mutations, which likely disrupt CTCF binding, have functional consequences that might be expected given important roles of CTCF as an architectural protein and ‘insulator’ for chromatin states.

In addition to CTCF sites, the binding sites for the ETS family of transcription factors (TFs) also register very high mutation rates. However, unlike CTCF sites, this effect is observed in skin

cancers but not in other cancer types [50,51]. Also in contrast to CTCF sites, differential DNA repair does not appear to play a role in the increased mutation rates at ETS sites. Instead, it was proposed that the binding of ETS TFs increases the propensity of DNA to be damaged by UV radiation [50,51]. Beyond the strong effect of bound CTCF and the ETS-family TFs, there is also a more modest but general effect of TF binding (not particular to any specific TF) that increases mutation rates by a different mechanism -- by interfering with NER activity [52,53]. This is observable in skin cancer and in other cancers exposed to mutagens that can be countered by NER, such as lung cancers, but is absent in other common cancers such as colon or breast. One intriguing observation is that mutation rates tend to be more strongly increased at TF sites in promoters than at the same TF sites in enhancers [52]. This might hint at different manner of TF binding between promoter and enhancers, or, alternatively, may suggest that TF binding is statistically associated with mutability but not itself causal, with another factor at promoters being the cause of the local deficiency in NER. Caution must be exercised in interpreting the effects of individual TFs on mutation rates because various TFs often bind in clusters and the clusters are also often marked by CTCF/cohesin binding [54]. Deconvoluting the effect of a particular TF from the neighboring binding of CTCF and other factors is critical for measuring the effect size of individual contributors.

In addition to the binding of CTCF and certain transcription factors, nucleosome occupancy has long been associated with altered patterns of genetic variation in populations and across species [55–59]. Consistent with this, nucleosome occupancy was shown to associate with subtle local changes in somatic mutation rates, evident in a periodic pattern corresponding to the internucleosomal distance of approximately 200 nt and additionally an approximately 10nt periodicity corresponding to the rotational constraints of DNA wrapped around nucleosomes [60–63]. Several lines of evidence suggest that this is due to differential DNA repair (there is evidence for MMR, NER and possibly BER playing a role): the pattern appears dependant upon functional repair pathways and it is evident in mutational signatures associated with repair deficiencies. Additionally, there is evidence that some types of DNA damage, in particular cytosine deamination and UV-catalyzed pyrimidine dimer formation, is modulated by nucleosome binding [55,60].

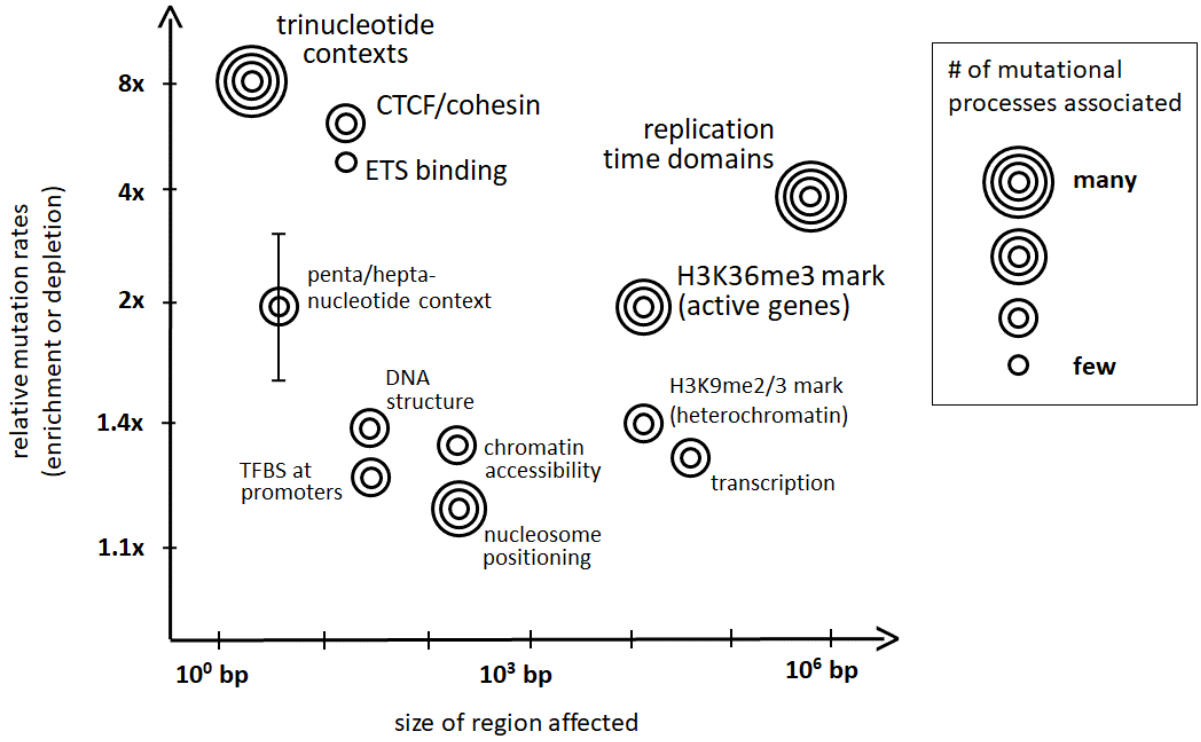

Figure 2. Associations of regional mutation rates with genomic and epigenomic variables, sorted by approximate size of affected region (*x* axis) and magnitude of enrichment/depletion (*y* axis).

Variability at the motif scale (1 - 10^1 bp). At the smallest scale, mutation rate at individual nucleotides can depend strongly on their immediate 5' and 3' neighbors, with these dependencies suggesting the mutational process that generated the mutation. For example, a C>T change in a TCW context (where W=A or T) commonly results from the activity of an enzyme from the APOBEC family of cytosine deaminases [64,65]. However the same C>T change but in an NCG context is commonly due to spontaneous cytosine deamination at methylated cytosines [6,7], while cytosines bearing the 5-hydroxymethyl modification were associated with more C>G changes but less C>T changes [7,66]. The relative abundance of mutations in different contexts informs about which mutagenic processes are ongoing in that individual. Conversely, analyzing how the mutability of trinucleotides changes across individuals identifies 'mutational signatures', mathematical constructs that aim to represent individual mutagenic mechanisms. More than 50 such signatures were found thus far, where mechanisms are known or proposed for some, and are quite varied, involving failures in DNA repair, exposure to mutagenic agents (exogenous or endogenous), and to spontaneous DNA damage

[36,61,67]. The mechanisms are, however, still unknown for many of these signatures or are speculative at best. It is also in many cases unclear that there is a one-to-one mapping of any particular signature to a mutagenic process, given that slightly different computational methods result in quite different sets of signatures. These issues are expected to lessen as the statistical methodologies mature.

Several extensions of the above mutational signatures framework are promising, in terms of better deconvoluting mutational processes from genomic data, and also in terms of helping interpret the processes and assigning them to a possible mechanism. Longer motifs including pentanucleotides and heptanucleotides have been examined for differential mutability in the soma and germline, establishing that these extended oligonucleotide contexts indeed matter for mutability in human for certain types of mutational processes [68–70]. However, for analysis of motifs longer than trinucleotides, the amount of data becomes limiting; methods have been proposed to ameliorate this [71]. Additionally, repetitive DNA motifs with propensity to form certain DNA structures were found to be associated with certain types of mutations [72,73], suggesting that DNA structure might be causal to mutagenesis. The difficulty of calling mutations at repetitive DNA from short-read sequencing are a challenge for such analyses. In addition to the most common mutation type -- single nucleotide substitutions -- mutational signatures were recently proposed for short insertions and deletions [74], for structural variants [75,76] and for clustered mutations [15]. Certain mutagenic mechanisms may be predominant in one type of mutational signature, for instance mutagenic activity due to use of error-prone DNA polymerase eta appears widespread and detectable *via* patterns of clustered mutations [15,77]. In other instances, the same process is evident across different types of mutational signatures. For instance, MMR deficiencies result in a mutation spectrum rich in e.g. GCN>GTN changes (among others) and additionally they result in an increased burden of indels, typically with a bias towards short deletions. Similarly, deficiency in the homologous recombination pathway results in both single-nucleotide changes of a certain spectrum [36,78], in short deletions with microhomology at borders, and in a pattern of copy-number changes [75,79] .

The redistribution of mutations in cancer

It is well expected that exposures to DNA damaging agents and failing DNA repair increase overall mutation rates. However genome sequencing of cancers has also provided evidence of another, less appreciated, but similarly widespread phenomenon: that exposure to mutagens and DNA repair failures also cause changes in the relative mutation rates of chromosomal

regions. Such 'redistribution' of mutations across the genome due to mutagenic exposures is likely to have important functional consequences when it occurs, in particular it may often increase the burden of mutations specifically in the regions where mutations have higher impact. This is because the default mutation rate landscape in mammalian genomes implies that active genes are protected *via* a combination of factors that preferentially recruit DNA repair: early DNA replication, transcription, chromatin accessibility, and the H3K36me3 chromatin mark. This protection of active chromatin by focused DNA repair is known to be lost in some mutator phenotypes and upon exposure to carcinogens, as we outline below.

Very prominently, human and other eukaryotic cells that lose MMR activity also lose the protection of early-replicating domains against mutation [8]. Not only does MMR loss imply an overall increase in mutation rates, but the increase is proportionally much larger at early-replicating, euchromatic, gene-rich domains. This is in contrast to another type of mutator phenotype due to mutations in the proofreading domain of DNA polymerase epsilon: these cancers usually have an even higher mutation burden than MMR-deficient cancers however they retain protection of early-replicating domains [8].

In addition to global MMR failure, other types of DNA repair alterations were linked with mutation redistribution. For instance, bladder cancers with mutations in the ERCC2 gene, encoding a helicase that participates in the NER pathway, show an altered mutation trinucleotide spectrum [80,81] but also show a loss of mutation protection in active chromatin.

Another, more widespread instance of mutation redistribution in many cancer types was recently proposed to occur via activity of error-prone, non-canonical MMR pathway [15]. This pathway is known to use the TLS DNA polymerase eta (a product of the POLH gene, also called XPV) during MMR [82,83], and the mutational pattern of clustered mutations at A:T nucleotide pairs observed in many tumors is consistent with activity of POLH [77]. Of note, while clustered mutations are rather rare, it is likely that POLH also creates a large number of (unclustered) single-nucleotide changes, which are directed towards H3K36me3-marked active genes, including cancer driver genes [15]. In multiple cancer types, this clustered mutational pattern is associated with exposure to various carcinogens, such as UV light, tobacco smoke and oxidative damage [15]; this is again consistent with the experimental observation that various mutagenic insults can recruit POLH to chromatin [82,83]. Interestingly, the mutational patterns of POLH were associated with alcohol consumption in multiple cancer types, providing a

possible mechanism by which alcohol is carcinogenic: rather than increasing the global mutation rate, alcohol or its metabolites increase the local mutation rate in active genes by triggering error-prone DNA synthesis [15]. Consistent with this, multiple other studies have reported an enrichment of mutations at A:T nucleotide pairs in esophageal and liver cancers of alcohol consumers [84,85] .

However, redistribution of mutations also happens not only due to inactivity (or non-canonical activity) of DNA repair pathways, but also because of exposure to certain DNA damaging agents. A salient examples of this is an endogenous mutagen -- the APOBEC family of cytosine deaminases, where the APOBEC3A and APOBEC3B paralogs are thought to cause mutations in cancer [64,86,87]. In contrast to most other mutation types, mutations due to APOBEC activity are enriched in early-replicating, gene-rich regions [88,89] and, consistently, have a high potential for generating impactful mutations in cancer driver genes [90,91].

Additional examples of mutations that are directed towards active genes have a less clear mechanistic basis but nonetheless may have a substantial mutagenic footprint (in all cases, shown for a rather restricted range of cell types): transcription-associated mutagenesis, with a strong DNA strand bias, reported in liver cancers [39], a modest increase in mutation burden at transcribed genes in healthy neurons [40], and a high burden of small indels in very highly expressed lineage-specific genes in lung, liver, thyroid and stomach tumors [41]. Finally, one mutagen that is known to target active, highly expressed regions (mostly gene promoters and downstream regions) is AID [92], a cytosine deaminase which acts in B-lymphocytes in a physiological (not pathological) process that serves to diversify antibody genes. It has been reported that AID can be ectopically expressed and cause mutations in non-lymphoid tissues [93–95]; indeed, many cancers of various types exhibit a clustered mutational signature apparently consistent with AID, which is also broadly targeted to active regions [15]. AID mutagenesis is a known cause of lymphoid cancers, and the current data, while only suggestive, merits further research to investigate a potential role in mutagenizing diverse non-blood tissues by AID or an AID-like activity, with potential for causing cancer therein.

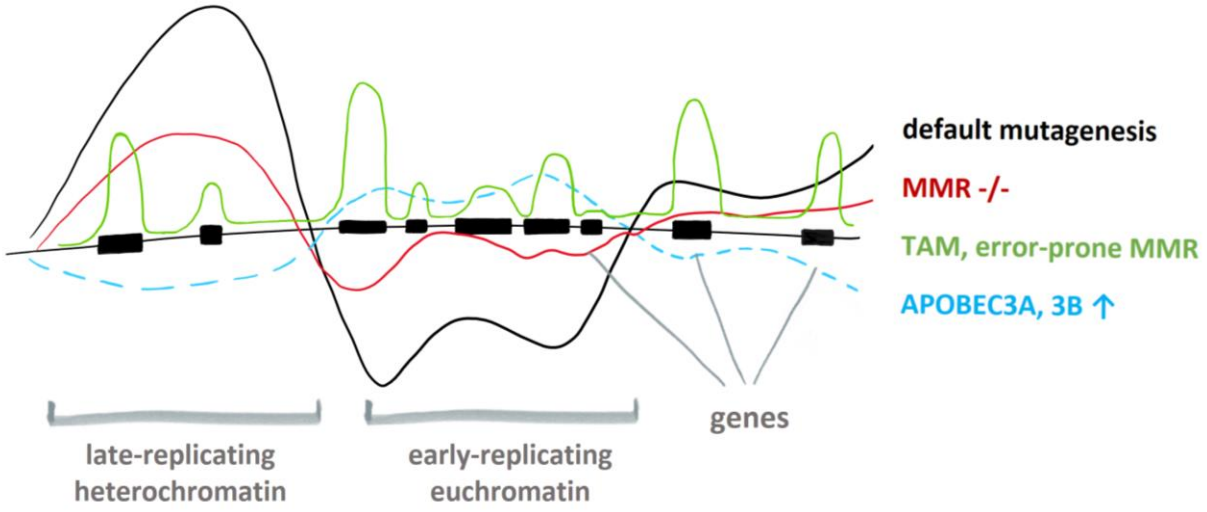

Figure 3. Redistribution of relative mutation rates towards active chromatin by failures of DNA repair and by exposure to mutagens. TAM, transcription-associated mutagenesis. MMR, DNA mismatch repair.

Outlook: challenges in genomic studies of mutation patterns.

Statistical analyses of mutation distributions across the genomes of somatic cells – including not only tumors but also healthy cells [96,97] and cell lines [26,98,99] – have provided valuable insights into the mechanisms that underlie mutation rate variation. The approaches that have been used can, however, likely still be improved to provide deeper insights into mechanisms of DNA repair and mutagenesis.

First, of course, there is a need for more data. A larger number of whole genome sequences will afford more statistical power to examine mutation rate variability at finer resolutions.

Second, there is a need to further develop and refine specialized statistical methods for discovering associations between mutation distributions and various genomic features, while controlling for both the strong confounding correlations that exist between many (epi)genomic features (such as chromatin accessibility, replication timing and transcription) and the heterogeneity of mutational processes across tumors (e.g. early replicating regions are G+C rich thus a process that mutates only C:G pairs creates an artefactual enrichment in early

replicating DNA). We have applied one statistical methodology to this problem [15,20] but it is likely that further developments and additional approaches will be required to fully address these issues. Adjusting for the known confounders increases the confidence that the discovered associations may be causal and thus inform about the mechanism behind the generation of (or protection from) mutations in certain regions.

However, in observational studies, no matter how carefully conducted, causality cannot be inferred with full certainty. Therefore, a third direction for research which is complementary to the analysis of cancer genomes is to perform controlled experiments, generating mutations by perturbing DNA repair pathways or chromatin modification pathways or by exposure to specific mutagens, and then observing the genome-wide patterns that emerge. This type of work has been recently initiated in multiple model systems including mice [20,100,101], mammalian and chicken cell lines [26,98,99,102], *C. elegans* [103,104], and yeast [5]. The genome-wide patterns observed in animal/cell-line models can then be matched to those observed in human samples, thus assigning a mechanistic basis for the mutational biases. This was performed mainly for trinucleotide mutation spectra but other kinds of patterns can, in principle, be examined in the same way.

The challenges in interpreting genome-wide studies of mutation rates highlighted above mean that it is often necessary to be cautious when interpreting the results of correlation-based studies. In particular, when a known (or unknown) confounding influence is not correctly controlled for, spurious associations can be reported. Moreover, while studies often focus attention on the statistical significance of the associations that they detect, attention should be paid to effect sizes. Here, in a broad sense, effect size could be understood to imply: the magnitude of the increase or decrease in mutation rates, or the target size (the number of nucleotides in the genome affected), or the number of mutational processes for which this pattern is relevant. Patterns which are subtle in magnitude (e.g. reduced mutation rates at accessible chromatin), which affect narrow regions of the genome (e.g. CTCF/cohesin bound sites), or which occur only under rarely occurring mutational regimes (e.g. hypermutation due to UV damage at TF binding sites) can be important because they provide insight into DNA repair and other molecular mechanisms. However, the impact of these biases on cells can be relatively small compared to other mechanisms. In contrast, the domain-scale mutation rate variability associated with replication time affects the entire genome, has large differences in rates between peaks and valleys, and is evident for many individual mutational processes.

A final note of caution concerns making analogies between mutation rate variation in somatic cells and the germline – the latter are, of course, also of high interest because they shape the genetic variability of populations, drive evolution, and impact the occurrence of genetic diseases. Mutation patterns have been examined in the human germline by analysis of population diversity [68,69,105], *de novo* mutations gleaned from sequencing parent-offspring trios [106,107] and also the evolutionary divergence between primates [17,18,59,108]. While indeed certain trends are, at first instance, similar to those observed in somatic cells, there are also substantial differences in mutation rate variability between the human germline and soma, including an overall reduced association of mutations with replication timing at the megabase scale [18,21,109]. These differences and the mechanisms underlying them are an important direction for future work.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This work was funded by the ERC Starting Grant “HYPER-INSIGHT” (to F.S.) and the ERC Consolidator Grant “IR-DC” (to B.L.). F.S. and B.L. are funded by the ICREA Research Professor programme. F.S. acknowledges support of the Severo Ochoa Centres of Excellence programme to the IRB Barcelona, and B.L. to CRG.

References

[1] L.A. Loeb, Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences, Cancer Res. 76 (2016) 2057–2059. doi:10.1158/0008-5472.CAN-16-0794.

[2] L.A. Loeb, Human cancers express mutator phenotypes: origin, consequences and targeting, Nat. Rev. Cancer. 11 (2011) 450–457. doi:10.1038/nrc3063.

[3] S.A. Roberts, D.A. Gordenin, Hypermutation in human cancer genomes: footprints and mechanisms, Nat. Rev. Cancer. 14 (2014) 786–800.

[4] K. Yi, Y.S. Ju, Patterns and mechanisms of structural variations in human cancer, Exp. Mol. Med. 50 (2018) 98. doi:10.1038/s12276-018-0112-3.

[5] S.A. Lujan, A.R. Clausen, A.B. Clark, H.K. MacAlpine, D.M. MacAlpine, E.P. Malc, P.A. Mieczkowski, A.B. Burkholder, D.C. Fargo, D.A. Gordenin, T.A. Kunkel, Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition, Genome Res. 24 (2014) 1751–1764. doi:10.1101/gr.178335.114.

[6] R.C. Poulos, J. Olivier, J.W.H. Wong, The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes, Nucleic Acids Res. 45 (2017) 7786–7795. doi:10.1093/nar/gkx463.

[7] M. Tomkova, M. McClellan, S. Kriaucionis, B. Schuster-Boeckler, 5-hydroxymethylcytosine marks regions with reduced mutation frequency in human DNA, ELife. 5 (2016) e17082. doi:10.7554/eLife.17082.

[8] F. Supek, B. Lehner, Differential DNA mismatch repair underlies mutation rate variation across the human genome, Nature. 521 (2015) 81–84. doi:10.1038/nature14173.

[9] S.R. Kennedy, L.A. Loeb, A.J. Herr, Somatic mutations in aging, cancer and neurodegeneration, Mech. Ageing Dev. 133 (2012) 118–126. doi:10.1016/j.mad.2011.10.009.

[10] V.B. Kaiser, M.S. Taylor, C.A. Semple, Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types, PLOS Genet. 12 (2016) e1006207. doi:10.1371/journal.pgen.1006207.

[11] R. Katainen, K. Dave, E. Pitkänen, K. Palin, T. Kivioja, N. Välimäki, A.E. Gylfe, H. Ristolainen, U.A. Hänninen, T. Cajuso, J. Kondelin, T. Tanskanen, J.-P. Mecklin, H. Järvinen, L. Renkonen-Sinisalo, A. Lepistö, E. Kaasinen, O. Kilpivaara, S. Tuupanen, M. Enge, J. Taipale, L.A. Aaltonen, CTCF/cohesin-binding sites are frequently mutated in cancer, Nat. Genet. 47 (2015) 818–821. doi:10.1038/ng.3335.

[12] R.C. Poulos, J.A.I. Thoms, Y.F. Guan, A. Unnikrishnan, J.E. Pimanda, J.W.H. Wong, Functional Mutations Form at CTCF-Cohesin Binding Sites in Melanoma Due to Uneven Nucleotide Excision Repair across the Motif, Cell Rep. 17 (2016) 2865–2872. doi:10.1016/j.celrep.2016.11.055.

[13] F. Li, G. Mao, D. Tong, J. Huang, L. Gu, W. Yang, G.-M. Li, The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through its Interaction with MutSα, Cell. 153 (2013) 590–600. doi:10.1016/j.cell.2013.03.025.

[14] S.X. Pfister, S. Ahrabi, L.-P. Zalmas, S. Sarkar, F. Aymard, C.Z. Bachrati, T. Helleday, G. Legube, N.B. La Thangue, A.C.G. Porter, T.C. Humphrey, SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability, Cell Rep. 7 (2014) 2006–2018. doi:10.1016/j.celrep.2014.05.026.

[15] F. Supek, B. Lehner, Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes, Cell. 170 (2017) 534-547.e23. doi:10.1016/j.cell.2017.07.003.

[16] K. Satou, M. Hori, K. Kawai, H. Kasai, H. Harashima, H. Kamiya, Involvement of specialized DNA polymerases in mutagenesis by 8-hydroxy-dGTP in human cells, DNA Repair. 8 (2009) 637–642. doi:10.1016/j.dnarep.2008.12.009.

[17] A. Hodgkinson, Y. Chen, A. Eyre-Walker, The large-scale distribution of somatic mutations in cancer genomes, Hum. Mutat. 33 (2012) 136–143. doi:10.1002/humu.21616.

[18] B. Schuster-Böckler, B. Lehner, Chromatin organization is a major influence on regional mutation rates in human cancer cells, Nature. 488 (2012) 504–507. doi:10.1038/nature11273.

[19] Y.H. Woo, W.-H. Li, DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes, Nat. Commun. 3 (2012) 1004. doi:10.1038/ncomms1982.

[20] A. Avgustinova, A. Symeonidi, A. Castellanos, U. Urdiroz-Urricelqui, L. Solé-Boldo, M. Martín, I. Pérez-Rodríguez, N. Prats, B. Lehner, F. Supek, S.A. Benitah, Loss of G9a preserves mutation patterns but increases chromatin accessibility, genomic instability and aggressiveness in skin tumours, Nat. Cell Biol. 20 (2018) 1400. doi:10.1038/s41556-018-0233-x.

[21] L. Liu, S. De, F. Michor, DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes, Nat. Commun. 4 (2013) 1502. doi:10.1038/ncomms2502.
[22] P. Polak, M.S. Lawrence, E. Haugen, N. Stoletzki, P. Stojanov, R.E. Thurman, L.A. Garraway, S. Mirkin, G. Getz, J.A. Stamatoyannopoulos, S.R. Sunyaev, Reduced local mutation density in regulatory DNA of cancer genomes is linked to DNA repair, Nat. Biotechnol. 32 (2014) 71–75. doi:10.1038/nbt.2778.
[23] Y. Gindin, M.S. Valenzuela, M.I. Aladjem, P.S. Meltzer, S. Bilke, A chromatin structure-based model accurately predicts DNA replication timing in human cells, Mol. Syst. Biol. 10 (2014). doi:10.1002/msb.134859.
[24] P. Polak, R. Karlić, A. Koren, R. Thurman, R. Sandstrom, M.S. Lawrence, A. Reynolds, E. Rynes, K. Vlahoviček, J.A. Stamatoyannopoulos, S.R. Sunyaev, Cell-of-origin chromatin organization shapes the mutational landscape of cancer, Nature. 518 (2015) 360–364. doi:10.1038/nature14221.
[25] J. Drost, R. van Boxtel, F. Blokzijl, T. Mizutani, N. Sasaki, V. Sasselli, J. de Ligt, S. Behjati, J.E. Grolleman, T. van Wezel, S. Nik-Zainal, R.P. Kuiper, E. Cuppen, H. Clevers, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer, Science. 358 (2017) 234–238. doi:10.1126/science.aao3130.
[26] X. Zou, M. Owusu, R. Harris, S.P. Jackson, J.I. Loizou, S. Nik-Zainal, Validating the concept of mutational signatures with isogenic cell models, Nat. Commun. 9 (2018) 1744. doi:10.1038/s41467-018-04052-8.
[27] M. Tomkova, J. Tomek, S. Kriaucionis, B. Schuster-Böckler, Mutational signature distribution varies with DNA replication timing and strand asymmetry, Genome Biol. 19 (2018) 129. doi:10.1186/s13059-018-1509-y.
[28] H. Zhao, B. Thienpont, B.T. Yesilyurt, M. Moisse, J. Reumers, L. Coenegrachts, X. Sagaert, S. Schrauwen, D. Smeets, G. Matthijs, S. Aerts, J. Cools, A. Metcalf, A. Spurdle, ANECS, F. Amant, D. Lambrechts, Mismatch repair deficiency endows tumors with a unique mutation signature and sensitivity to DNA double-strand breaks, ELife. 3 (2014) e02725. doi:10.7554/eLife.02725.
[29] J.D. Hawk, L. Stefanovic, J.C. Boyer, T.D. Petes, R.A. Farber, Variation in efficiency of DNA mismatch repair at different sites in the yeast genome, Proc. Natl. Acad. Sci. 102 (2005) 8639–8643. doi:10.1073/pnas.0503415102.
[30] E.J. Belfield, Z.J. Ding, F.J.C. Jamieson, A.M. Visscher, S.J. Zheng, A. Mithani, N.P. Harberd, DNA mismatch repair preferentially protects genes from mutation, Genome Res. 28 (2018) 66–74. doi:10.1101/gr.219303.116.
[31] C.L. Zheng, N.J. Wang, J. Chung, H. Moslehi, J.Z. Sanborn, J.S. Hur, E.A. Collisson, S.S. Vemula, A. Naujokas, K.E. Chiotti, J.B. Cheng, H. Fassihi, A.J. Blumberg, C.V. Bailey, G.M. Fudem, F.G. Mihm, B.B. Cunningham, I.M. Neuhaus, W. Liao, D.H. Oh, J.E. Cleaver, P.E. LeBoit, J.F. Costello, A.R. Lehmann, J.W. Gray, P.T. Spellman, S.T. Arron, N. Huh, E. Purdom, R.J. Cho, Transcription Restores DNA Repair to Heterochromatin, Determining Regional Mutation Rates in Cancer Genomes, Cell Rep. 9 (2014) 1228–1234. doi:10.1016/j.celrep.2014.10.031.
[32] M. Elez, A.W. Murray, L.-J. Bi, X.-E. Zhang, I. Matic, M. Radman, Seeing Mutations in Living Cells, Curr. Biol. 20 (2010) 1432–1437. doi:10.1016/j.cub.2010.06.071.
[33] H. Hombauer, C.S. Campbell, C.E. Smith, A. Desai, R.D. Kolodner, Visualization of Eukaryotic DNA Mismatch Repair Reveals Distinct Recognition and Repair Intermediates, Cell. 147 (2011) 1040–1053. doi:10.1016/j.cell.2011.10.025.
[34] Y. Huang, L. Gu, G.-M. Li, H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation, J. Biol. Chem. 293 (2018) 7811–7823. doi:10.1074/jbc.RA118.002839.

[35] L. Lv, F. Wang, X. Ma, Y. Yang, Z. Wang, H. Liu, X. Li, Z. Liu, T. Zhang, M. Huang, E.C. Friedberg, T.-S. Tang, C. Guo, Mismatch repair protein MSH2 regulates translesion DNA synthesis following exposure of cells to UV radiation, Nucleic Acids Res. (2013) gkt793. doi:10.1093/nar/gkt793.

[36] L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, S.A.J.R. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, S. Boyault, B. Burkhardt, A.P. Butler, C. Caldas, H.R. Davies, C. Desmedt, R. Eils, J.E. Eyfjörd, J.A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinsk, N. Jäger, D.T.W. Jones, D. Jones, S. Knappskog, M. Kool, S.R. Lakhani, C. López-Otín, S. Martin, N.C. Munshi, H. Nakamura, P.A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J.V. Pearson, X.S. Puente, K. Raine, M. Ramakrishna, A.L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T.N. Schumacher, P.N. Span, J.W. Teague, Y. Totoki, A.N.J. Tutt, R. Valdés-Mas, M.M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L.R. Yates, A.P.C.G. Initiative, I.B.C. Consortium, I.M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P.A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S.M. Grimmond, R. Siebert, E. Campo, T. Shibata, S.M. Pfister, P.J. Campbell, M.R. Stratton, Signatures of mutational processes in human cancer, Nature. 500 (2013) 415–421. doi:10.1038/nature12477.

[37] E.D. Pleasance, R.K. Cheetham, P.J. Stephens, D.J. McBride, S.J. Humphray, C.D. Greenman, I. Varela, M.-L. Lin, G.R. Ordóñez, G.R. Bignell, K. Ye, J. Alipaz, M.J. Bauer, D. Beare, A. Butler, R.J. Carter, L. Chen, A.J. Cox, S. Edkins, P.I. Kokko-Gonzales, N.A. Gormley, R.J. Grocock, C.D. Haudenschild, M.M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L.J. Mudie, Z. Ning, T. Royce, O.B. Schulz-Trieglaff, A. Spiridou, L.A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M.T. Ross, P.J. Campbell, D.R. Bentley, P.A. Futreal, M.R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome, Nature. 463 (2010) 191–196. doi:10.1038/nature08658.

[38] E.D. Pleasance, P.J. Stephens, S. O'Meara, D.J. McBride, A. Meynert, D. Jones, M.-L. Lin, D. Beare, K.W. Lau, C. Greenman, I. Varela, S. Nik-Zainal, H.R. Davies, G.R. Ordoñez, L.J. Mudie, C. Latimer, S. Edkins, L. Stebbings, L. Chen, M. Jia, C. Leroy, J. Marshall, A. Menzies, A. Butler, J.W. Teague, J. Mangion, Y.A. Sun, S.F. McLaughlin, H.E. Peckham, E.F. Tsung, G.L. Costa, C.C. Lee, J.D. Minna, A. Gazdar, E. Birney, M.D. Rhodes, K.J. McKernan, M.R. Stratton, P.A. Futreal, P.J. Campbell, A small-cell lung cancer genome with complex signatures of tobacco exposure, Nature. 463 (2010) 184–190. doi:10.1038/nature08629.

[39] N.J. Haradhvala, P. Polak, P. Stojanov, K.R. Covington, E. Shinbrot, J.M. Hess, E. Rheinbay, J. Kim, Y.E. Maruvka, L.Z. Braunstein, A. Kamburov, P.C. Hanawalt, D.A. Wheeler, A. Koren, M.S. Lawrence, G. Getz, Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair, Cell. 164 (2016) 538–549. doi:10.1016/j.cell.2015.12.050.

[40] M.A. Lodato, M.B. Woodworth, S. Lee, G.D. Evrony, B.K. Mehta, A. Karger, S. Lee, T.W. Chittenden, A.M. D'Gama, X. Cai, L.J. Luquette, E. Lee, P.J. Park, C.A. Walsh, Somatic mutation in single human neurons tracks developmental and transcriptional history, Science. 350 (2015) 94–98. doi:10.1126/science.aab1785.

[41] M. Imielinski, G. Guo, M. Meyerson, Insertions and deletions target lineage-defining genes in human cancers, Cell. 168 (2017) 460-472.e14. doi:10.1016/j.cell.2016.12.025.

[42] P. Sen, W. Dang, G. Donahue, J. Dai, J. Dorsey, X. Cao, W. Liu, K. Cao, R. Perry, J.Y. Lee, B.M. Wasko, D.T. Carr, C. He, B. Robison, J. Wagner, B.D. Gregory, M. Kaeberlein, B.K. Kennedy, J.D. Boeke, S.L. Berger, H3K36 methylation promotes longevity by enhancing transcriptional fidelity, Genes Dev. 29 (2015) 1362–1376. doi:10.1101/gad.263707.115.

[43] T. Vavouri, B. Lehner, Human genes with CpG island promoters have a distinct transcription-associated chromatin organization, Genome Biol. 13 (2012) R110. doi:10.1186/gb-2012-13-11-r110.
[44] J. Frigola, R. Sabarinathan, L. Mularoni, F. Muiños, A. Gonzalez-Perez, N. López-Bigas, Reduced mutation rate in exons due to differential mismatch repair, Nat. Genet. 49 (2017) 1684–1692. doi:10.1038/ng.3991.
[45] C. Melton, J.A. Reuter, D.V. Spacek, M. Snyder, Recurrent Somatic Mutations in Regulatory Regions of Human Cancer Genomes, Nat. Genet. 47 (2015) 710–716. doi:10.1038/ng.3332.
[46] L.D. Hurst, N.N. Batada, Depletion of somatic mutations in splicing-associated sequences in cancer genomes, Genome Biol. 18 (2017) 213. doi:10.1186/s13059-017-1337-5.
[47] A.R. Poetsch, S.J. Boulton, N.M. Luscombe, Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis, Genome Biol. 19 (2018) 215. doi:10.1186/s13059-018-1582-2.
[48] P.E. García-Nieto, E.K. Schwartz, D.A. King, J. Paulsen, P. Collas, R.E. Herrera, A.J. Morrison, Carcinogen susceptibility is regulated by genome architecture and predicts cancer mutagenesis, EMBO J. 36 (2017) 2829–2843. doi:10.15252/embj.201796717.
[49] Y.A. Guo, M.M. Chang, W. Huang, W.F. Ooi, M. Xing, P. Tan, A.J. Skanderup, Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers, Nat. Commun. 9 (2018). doi:10.1038/s41467-018-03828-2.
[50] K. Elliott, M. Boström, S. Filges, M. Lindberg, J.V. den Eynden, A. Ståhlberg, A.R. Clausen, E. Larsson, Elevated pyrimidine dimer formation at distinct genomic bases underlies promoter mutation hotspots in UV-exposed cancers, PLOS Genet. 14 (2018) e1007849. doi:10.1371/journal.pgen.1007849.
[51] P. Mao, A.J. Brown, S. Esaki, S. Lockwood, G.M.K. Poon, M.J. Smerdon, S.A. Roberts, J.J. Wyrick, ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma, Nat. Commun. 9 (2018) 2626. doi:10.1038/s41467-018-05064-0.
[52] D. Perera, R.C. Poulos, A. Shah, D. Beck, J.E. Pimanda, J.W.H. Wong, Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes, Nature. 532 (2016) 259–263. doi:10.1038/nature17437.
[53] R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. López-Bigas, Nucleotide excision repair is impaired by binding of transcription factors to DNA, Nature. 532 (2016) 264–267. doi:10.1038/nature17661.
[54] J. Yan, M. Enge, T. Whitington, K. Dave, J. Liu, I. Sur, B. Schmierer, A. Jolma, T. Kivioja, M. Taipale, J. Taipale, Transcription Factor Binding in Human Cells Occurs in Dense Clusters Formed around Cohesin Anchor Sites, Cell. 154 (2013) 801–813. doi:10.1016/j.cell.2013.07.034.
[55] X. Chen, Z. Chen, H. Chen, Z. Su, J. Yang, F. Lin, S. Shi, X. He, Nucleosomes Suppress Spontaneous Mutations Base-Specifically in Eukaryotes, Science. 335 (2012) 1235–1238. doi:10.1126/science.1217580.
[56] S. Sasaki, C.C. Mello, A. Shimada, Y. Nakatani, S. Hashimoto, M. Ogawa, K. Matsushima, S.G. Gu, M. Kasahara, B. Ahsan, A. Sasaki, T. Saito, Y. Suzuki, S. Sugano, Y. Kohara, H. Takeda, A. Fire, S. Morishita, Chromatin-Associated Periodicity in Genetic Variation Downstream of Transcriptional Start Sites, Science. 323 (2009) 401–404. doi:10.1126/science.1163183.
[57] M.Y. Tolstorukov, N. Volfovsky, R.M. Stephens, P.J. Park, Impact of chromatin structure on sequence variability in the human genome, Nat. Struct. Mol. Biol. 18 (2011) 510–515. doi:10.1038/nsmb.2012.

[58] T. Warnecke, N.N. Batada, L.D. Hurst, The Impact of the Nucleosome Code on Protein-Coding Sequence Evolution in Yeast, PLOS Genet. 4 (2008) e1000250. doi:10.1371/journal.pgen.1000250.
[59] H. Ying, J. Epps, R. Williams, G. Huttley, Evidence that Localized Variation in Primate Sequence Divergence Arises from an Influence of Nucleosome Placement on DNA Repair, Mol. Biol. Evol. 27 (2010) 637–649. doi:10.1093/molbev/msp253.
[60] A.J. Brown, P. Mao, M.J. Smerdon, J.J. Wyrick, S.A. Roberts, Nucleosome positions establish an extended mutation signature in melanoma, PLOS Genet. 14 (2018) e1007823. doi:10.1371/journal.pgen.1007823.
[61] S. Morganella, L.B. Alexandrov, D. Glodzik, X. Zou, H. Davies, J. Staaf, A.M. Sieuwerts, A.B. Brinkman, S. Martin, M. Ramakrishna, A. Butler, H.-Y. Kim, Å. Borg, C. Sotiriou, P.A. Futreal, P.J. Campbell, P.N. Span, S. Van Laere, S.R. Lakhani, J.E. Eyfjord, A.M. Thompson, H.G. Stunnenberg, M.J. van de Vijver, J.W.M. Martens, A.-L. Børresen-Dale, A.L. Richardson, G. Kong, G. Thomas, J. Sale, C. Rada, M.R. Stratton, E. Birney, S. Nik-Zainal, The topography of mutational processes in breast cancer genomes, Nat. Commun. 7 (2016) 11383. doi:10.1038/ncomms11383.
[62] O. Pich, F. Muiños, R. Sabarinathan, I. Reyes-Salazar, A. Gonzalez-Perez, N. Lopez-Bigas, Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes, Cell. 175 (2018) 1074-1087.e18. doi:10.1016/j.cell.2018.10.004.
[63] P.G. Yazdi, B.A. Pedersen, J.F. Taylor, O.S. Khattab, Y.-H. Chen, Y. Chen, S.E. Jacobsen, P.H. Wang, Increasing Nucleosome Occupancy Is Correlated with an Increasing Mutation Rate so Long as DNA Repair Machinery Is Intact, PLoS ONE. 10 (2015) e0136574. doi:10.1371/journal.pone.0136574.
[64] M.B. Burns, N.A. Temiz, R.S. Harris, Evidence for APOBEC3B mutagenesis in multiple human cancers, Nat. Genet. 45 (2013) 977–983. doi:10.1038/ng.2701.
[65] S.A. Roberts, M.S. Lawrence, L.J. Klimczak, S.A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, G.V. Kryukov, S.L. Carter, G. Saksena, S. Harris, R.R. Shah, M.A. Resnick, G. Getz, D.A. Gordenin, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers, Nat. Genet. 45 (2013) 970–976. doi:10.1038/ng.2702.
[66] F. Supek, B. Lehner, P. Hajkova, T. Warnecke, Hydroxymethylated Cytosines Are Associated with Elevated C to G Transversion Rates, PLoS Genet. 10 (2014) e1004585. doi:10.1371/journal.pgen.1004585.
[67] L.B. Alexandrov, P.H. Jones, D.C. Wedge, J.E. Sale, P.J. Campbell, S. Nik-Zainal, M.R. Stratton, Clock-like mutational processes in human somatic cells, Nat. Genet. 47 (2015) 1402–1407. doi:10.1038/ng.3441.
[68] V. Aggarwala, B.F. Voight, An expanded sequence context model broadly explains variability in polymorphism levels across the human genome, Nat. Genet. 48 (2016) 349–355. doi:10.1038/ng.3511.
[69] J. Carlson, A.E. Locke, M. Flickinger, M. Zawistowski, S. Levy, R.M. Myers, M. Boehnke, H.M. Kang, L.J. Scott, J.Z. Li, S. Zöllner, Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans, Nat. Commun. 9 (2018) 3753. doi:10.1038/s41467-018-05936-5.
[70] I. Martincorena, K.M. Raine, M. Gerstung, K.J. Dawson, K. Haase, P.V. Loo, H. Davies, M.R. Stratton, P.J. Campbell, Universal Patterns of Selection in Cancer and Somatic Tissues, Cell. 171 (2017) 1029-1041.e21. doi:10.1016/j.cell.2017.09.042.
[71] Y. Shiraishi, G. Tremmel, S. Miyano, M. Stephens, A Simple Model-Based Approach to Inferring and Visualizing Cancer Mutation Signatures, PLoS Genet. 11 (2015) e1005657. doi:10.1371/journal.pgen.1005657.

[72] C. Duan, Q. Huan, X. Chen, S. Wu, L.B. Carey, X. He, W. Qian, Reduced intrinsic DNA curvature leads to increased mutation rate, Genome Biol. 19 (2018) 132. doi:10.1186/s13059-018-1525-y.

[73] I. Georgakopoulos-Soares, S. Morganella, N. Jain, M. Hemberg, S. Nik-Zainal, Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis, Genome Res. 28 (2018) 1264–1271. doi:10.1101/gr.231688.117.

[74] L.B. Alexandrov, J. Kim, N.J. Haradhvala, M.N. Huang, A.W. Ng, A. Boot, K.R. Covington, D.A. Gordenin, E. Bergstrom, N. Lopez-Bigas, L.J. Klimczak, J.R. McPherson, S. Morganella, R. Sabarinathan, D.A. Wheeler, V. Mustonen, G. Getz, S.G. Rozen, M.R. Stratton, on behalf of the P.M.S.W.G. and the I.P.-C.A. of W.G. Network, The Repertoire of Mutational Signatures in Human Cancer, BioRxiv. (2018) 322859. doi:10.1101/322859.

[75] G. Macintyre, T.E. Goranova, D. De Silva, D. Ennis, A.M. Piskorz, M. Eldridge, D. Sie, L.-A. Lewsley, A. Hanif, C. Wilson, S. Dowson, R.M. Glasspool, M. Lockley, E. Brockbank, A. Montes, A. Walther, S. Sundar, R. Edmondson, G.D. Hall, A. Clamp, C. Gourley, M. Hall, C. Fotopoulou, H. Gabra, J. Paul, A. Supernat, D. Millan, A. Hoyle, G. Bryson, C. Nourse, L. Mincarelli, L. Navarro Sanchez, B. Ylstra, M. Jimenez-Linan, L. Moore, O. Hofmann, F. Markowetz, I.A. McNeish, J.D. Brenton, Copy-number signatures and mutational processes in ovarian carcinoma, Nat. Genet. 50 (2018) 1262–1270. doi:10.1038/s41588-018-0179-8.

[76] S. Nik-Zainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L.B. Alexandrov, S. Martin, D.C. Wedge, P. Van Loo, Y.S. Ju, M. Smid, A.B. Brinkman, S. Morganella, M.R. Aure, O.C. Lingjærde, A. Langerød, M. Ringnér, S.-M. Ahn, S. Boyault, J.E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J.A. Foekens, M. Gerstung, G.K.J. Hooijer, S.J. Jang, D.R. Jones, H.-Y. Kim, T.A. King, S. Krishnamurthy, H.J. Lee, J.-Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C.A. Purdie, K. Raine, K. Ramakrishnan, F.G. Rodríguez-González, G. Romieu, A.M. Sieuwerts, P.T. Simpson, R. Shepherd, L. Stebbings, O.A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G.G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B.K.T. Tan, J. Jonkers, Å. Borg, N.T. Ueno, C. Sotiriou, A. Viari, P.A. Futreal, P.J. Campbell, P.N. Span, S. Van Laere, S.R. Lakhani, J.E. Eyfjord, A.M. Thompson, E. Birney, H.G. Stunnenberg, M.J. van de Vijver, J.W.M. Martens, A.-L. Børresen-Dale, A.L. Richardson, G. Kong, G. Thomas, M.R. Stratton, Landscape of somatic mutations in 560 breast cancer whole-genome sequences, Nature. 534 (2016) 47–54. doi:10.1038/nature17676.

[77] I.B. Rogozin, A. Goncearenco, A.G. Lada, S. De, V. Yurchenko, G. Nudelman, A.R. Panchenko, D.N. Cooper, Y.I. Pavlov, DNA polymerase η mutational signatures are found in a variety of different types of cancer, Cell Cycle. 17 (2018) 348–355. doi:10.1080/15384101.2017.1404208.

[78] P. Polak, J. Kim, L.Z. Braunstein, R. Karlic, N.J. Haradhavala, G. Tiao, D. Rosebrock, D. Livitz, K. Kübler, K.W. Mouw, A. Kamburov, Y.E. Maruvka, I. Leshchiner, E.S. Lander, T.R. Golub, A. Zick, A. Orthwein, M.S. Lawrence, R.N. Batra, C. Caldas, D.A. Haber, P.W. Laird, H. Shen, L.W. Ellisen, A.D. D'Andrea, S.J. Chanock, W.D. Foulkes, G. Getz, A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer, Nat. Genet. 49 (2017) 1476–1486. doi:10.1038/ng.3934.

[79] F. Menghi, F.P. Barthel, V. Yadav, M. Tang, B. Ji, Z. Tang, G.W. Carter, Y. Ruan, R. Scully, R.G.W. Verhaak, J. Jonkers, E.T. Liu, The Tandem Duplicator Phenotype Is a Prevalent Genome-Wide Cancer Configuration Driven by Distinct Gene Mutations, Cancer Cell. 34 (2018) 197-210.e5. doi:10.1016/j.ccell.2018.06.008.

[80] G. Guo, X. Sun, C. Chen, S. Wu, P. Huang, Z. Li, M. Dean, Y. Huang, W. Jia, Q. Zhou, A. Tang, Z. Yang, X. Li, P. Song, X. Zhao, R. Ye, S. Zhang, Z. Lin, M. Qi, S. Wan, L. Xie, F. Fan, M.L. Nickerson, X. Zou, X. Hu, L. Xing, Z. Lv, H. Mei, S. Gao, C. Liang, Z. Gao, J.

Lu, Y. Yu, C. Liu, L. Li, X. Fang, Z. Jiang, J. Yang, C. Li, X. Zhao, J. Chen, F. Zhang, Y. Lai, Z. Lin, F. Zhou, H. Chen, H.C. Chan, S. Tsang, D. Theodorescu, Y. Li, X. Zhang, J. Wang, H. Yang, Y. Gui, J. Wang, Z. Cai, Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation, Nat. Genet. 45 (2013) 1459–1463. doi:10.1038/ng.2798.

[81] J. Kim, K.W. Mouw, P. Polak, L.Z. Braunstein, A. Kamburov, G. Tiao, D.J. Kwiatkowski, J.E. Rosenberg, E.M. Van Allen, A.D. D'Andrea, G. Getz, Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors, Nat. Genet. 48 (2016) 600–606. doi:10.1038/ng.3557.

[82] J. Peña-Diaz, S. Bregenhorn, M. Ghodgaonkar, C. Follonier, M. Artola-Borán, D. Castor, M. Lopes, A.A. Sartori, J. Jiricny, Noncanonical Mismatch Repair as a Source of Genomic Instability in Human Cells, Mol. Cell. 47 (2012) 669–680. doi:10.1016/j.molcel.2012.07.006.

[83] A. Zlatanou, E. Despras, T. Braz-Petta, I. Boubakour-Azzouz, C. Pouvelle, G.S. Stewart, S. Nakajima, A. Yasui, A.A. Ishchenko, P.L. Kannouche, The hMsh2-hMsh6 complex acts in concert with monoubiquitinated PCNA and Pol η in response to oxidative DNA damage in human cells, Mol. Cell. 43 (2011) 649–662. doi:10.1016/j.molcel.2011.06.023.

[84] E. Letouzé, J. Shinde, V. Renault, G. Couchy, J.-F. Blanc, E. Tubacher, Q. Bayard, D. Bacq, V. Meyer, J. Semhoun, P. Bioulac-Sage, S. Prévôt, D. Azoulay, V. Paradis, S. Imbeaud, J.-F. Deleuze, J. Zucman-Rossi, Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis, Nat. Commun. 8 (2017). doi:10.1038/s41467-017-01358-x.

[85] X.C. Li, M.Y. Wang, M. Yang, H.J. Dai, B.F. Zhang, W. Wang, X.L. Chu, X. Wang, H. Zheng, R.F. Niu, W. Zhang, K.X. Chen, A mutational signature associated with alcohol consumption and prognostically significantly mutated driver genes in esophageal squamous cell carcinoma, Ann. Oncol. 29 (2018) 938–944. doi:10.1093/annonc/mdy011.

[86] V. Caval, R. Suspène, M. Shapira, J.-P. Vartanian, S. Wain-Hobson, A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′UTR enhances chromosomal DNA damage, Nat. Commun. 5 (2014) ncomms6129. doi:10.1038/ncomms6129.

[87] K. Chan, S.A. Roberts, L.J. Klimczak, J.F. Sterling, N. Saini, E.P. Malc, J. Kim, D.J. Kwiatkowski, D.C. Fargo, P.A. Mieczkowski, G. Getz, D.A. Gordenin, An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers, Nat. Genet. 47 (2015) 1067–1072. doi:10.1038/ng.3378.

[88] M.D. Kazanov, S.A. Roberts, P. Polak, J. Stamatoyannopoulos, L.J. Klimczak, D.A. Gordenin, S.R. Sunyaev, APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions, Cell Rep. 13 (2015) P1103-1109. doi:10.1016/j.celrep.2015.09.077.

[89] V.B. Seplyarskiy, R.A. Soldatov, K.Y. Popadin, S.E. Antonarakis, G.A. Bazykin, S.I. Nikolaev, APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication, Genome Res. 26 (2016) 174–182. doi:10.1101/gr.197046.115.

[90] V.L. Cannataro, S.G. Gaffney, T. Sasaki, N. Issaeva, N.K.S. Grewal, J.R. Grandis, W.G. Yarbrough, B. Burtness, K.S. Anderson, J.P. Townsend, APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma, Oncogene. (2019) 1. doi:10.1038/s41388-018-0657-6.

[91] D. Temko, I.P.M. Tomlinson, S. Severini, B. Schuster-Böckler, T.A. Graham, The effects of mutational processes and selection on driver mutations across cancer types, Nat. Commun. 9 (2018) 1857. doi:10.1038/s41467-018-04208-6.

[92] Q. Wang, T. Oliveira, M. Jankovic, I.T. Silva, O. Hakim, K. Yao, A. Gazumyan, C.T. Mayer, R. Pavri, R. Casellas, M.C. Nussenzweig, D.F. Robbiani, Epigenetic targeting of activation-induced cytidine deaminase, Proc. Natl. Acad. Sci. 111 (2014) 18667–18672. doi:10.1073/pnas.1420575111.

[93] J. Komori, H. Marusawa, T. Machimoto, Y. Endo, K. Kinoshita, T. Kou, H. Haga, I. Ikai, S. Uemoto, T. Chiba, Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma, Hepatology. 47 (2008) 888–896. doi:10.1002/hep.22125.

[94] T. Nonaka, Y. Toda, H. Hiai, M. Uemura, M. Nakamura, N. Yamamoto, R. Asato, Y. Hattori, K. Bessho, N. Minato, K. Kinoshita, Involvement of activation-induced cytidine deaminase in skin cancer development, J. Clin. Invest. 126 (2016) 1367–1382. doi:10.1172/JCI81522.

[95] T. Shimizu, H. Marusawa, Y. Matsumoto, T. Inuzuka, A. Ikeda, Y. Fujii, S. Minamiguchi, S. Miyamoto, T. Kou, Y. Sakai, J.E. Crabtree, T. Chiba, Accumulation of somatic mutations in TP53 in gastric epithelium with Helicobacter pylori infection, Gastroenterology. 147 (2014) 407-417.e3. doi:10.1053/j.gastro.2014.04.036.

[96] F. Blokzijl, J. de Ligt, M. Jager, V. Sasselli, S. Roerink, N. Sasaki, M. Huch, S. Boymans, E. Kuijk, P. Prins, I.J. Nijman, I. Martincorena, M. Mokry, C.L. Wiegerinck, S. Middendorp, T. Sato, G. Schwank, E.E.S. Nieuwenhuis, M.M.A. Verstegen, L.J.W. van der Laan, J. de Jonge, J.N.M. IJzermans, R.G. Vries, M. van de Wetering, M.R. Stratton, H. Clevers, E. Cuppen, R. van Boxtel, Tissue-specific mutation accumulation in human adult stem cells during life, Nature. 538 (2016) 260–264. doi:10.1038/nature19768.

[97] I. Franco, A. Johansson, K. Olsson, P. Vrtačnik, P. Lundin, H.T. Helgadottir, M. Larsson, G. Revêchon, C. Bosia, A. Pagnani, P. Provero, T. Gustafsson, H. Fischer, M. Eriksson, Somatic mutagenesis in satellite cells associates with human skeletal muscle aging, Nat. Commun. 9 (2018) 800. doi:10.1038/s41467-018-03244-6.

[98] M. Olivier, A. Weninger, M. Ardin, H. Huskova, X. Castells, M.P. Vallée, J. McKay, T. Nedelko, K.-R. Muehlbauer, H. Marusawa, J. Alexander, L. Hazelwood, G. Byrnes, M. Hollstein, J. Zavadil, Modelling mutational landscapes of human cancers in vitro, Sci. Rep. 4 (2014). doi:10.1038/srep04482.

[99] J. Zámborszky, B. Szikriszt, J.Z. Gervai, O. Pipek, Á. Póti, M. Krzystanek, D. Ribli, J.M. Szalai-Gindl, I. Csabai, Z. Szallasi, C. Swanton, A.L. Richardson, D. Szüts, Loss of BRCA1 or BRCA2 markedly increases the rate of base substitution mutagenesis and has distinct effects on genomic deletions, Oncogene. 36 (2017) 746–755. doi:10.1038/onc.2016.243.

[100] F. Connor, T.F. Rayner, S.J. Aitken, C. Feig, M. Lukk, J. Santoyo-Lopez, D.T. Odom, Mutational landscape of a chemically-induced mouse model of liver cancer, J. Hepatol. 69 (2018) 840–850. doi:10.1016/j.jhep.2018.06.009.

[101] D. Nassar, M. Latil, B. Boeckx, D. Lambrechts, C. Blanpain, Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma, Nat. Med. 21 (2015) 946–954. doi:10.1038/nm.3878.

[102] J.E. Kucab, X. Zou, S. Morganella, M. Joel, A.S. Nanda, E. Nagy, C. Gomez, A. Degasperi, R. Harris, S.P. Jackson, V.M. Arlt, D.H. Phillips, S. Nik-Zainal, A Compendium of Mutational Signatures of Environmental Agents, Cell. (2019). doi:10.1016/j.cell.2019.03.001.

[103] B. Meier, N.V. Volkova, Y. Hong, P. Schofield, P.J. Campbell, M. Gerstung, A. Gartner, Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers, Genome Res. 28 (2018) 666–675. doi:10.1101/gr.226845.117.

[104] B. Meier, S.L. Cooke, J. Weiss, A.P. Bailly, L.B. Alexandrov, J. Marshall, K. Raine, M. Maddison, E. Anderson, M.R. Stratton, A. Gartner, P.J. Campbell, C. elegans whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency, Genome Res. 24 (2014) 1624–1636. doi:10.1101/gr.175547.114.

[105] K. Harris, J.K. Pritchard, Rapid evolution of the human mutation spectrum, ELife. 6 (2017) e24284. doi:10.7554/eLife.24284.
[106] L.C. Francioli, P.P. Polak, A. Koren, A. Menelaou, S. Chun, I. Renkens, Genome of the Netherlands Consortium, C.M. van Duijn, M. Swertz, C. Wijmenga, G. van Ommen, P.E. Slagboom, D.I. Boomsma, K. Ye, V. Guryev, P.F. Arndt, W.P. Kloosterman, P.I.W. de Bakker, S.R. Sunyaev, Genome-wide patterns and properties of de novo mutations in humans, Nat. Genet. 47 (2015) 822–826. doi:10.1038/ng.3292.
[107] J.M. Goldmann, W.S.W. Wong, M. Pinelli, T. Farrah, D. Bodian, A.B. Stittrich, G. Glusman, L.E.L.M. Vissers, A. Hoischen, J.C. Roach, J.G. Vockley, J.A. Veltman, B.D. Solomon, C. Gilissen, J.E. Niederhuber, Parent-of-origin-specific signatures of *de novo* mutations, Nat. Genet. 48 (2016) 935–939. doi:10.1038/ng.3597.
[108] G. Ananda, F. Chiaromonte, K.D. Makova, A genome-wide view of mutation rate co-variation using multivariate analyses, Genome Biol. 12 (2011) R27. doi:10.1186/gb-2011-12-3-r27.
[109] C. Chen, H. Qi, Y. Shen, J. Pickrell, M. Przeworski, Contrasting Determinants of Mutation Rates in Germline and Soma, Genetics. 207 (2017) 255–267. doi:10.1534/genetics.117.1114.